YOLACT
Real-time Instance Segmentation

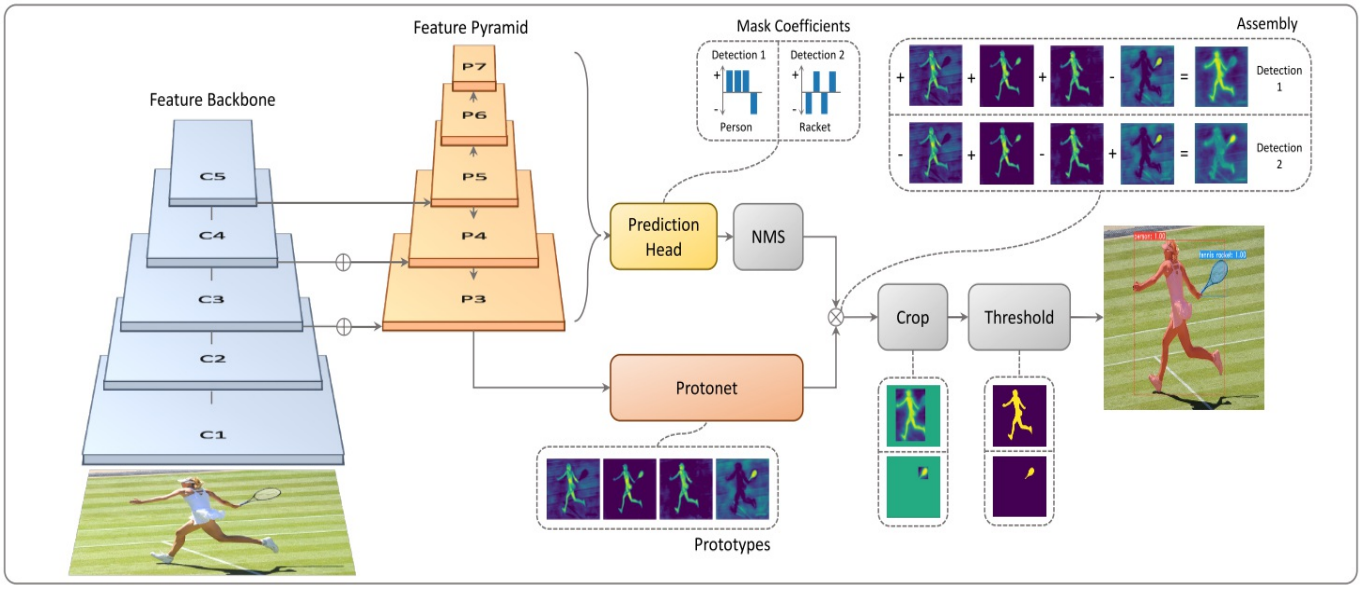
Daniel Bolya Chong Zhou Fanyi Xiao Yong Jae Lee
University of California, Davis
{dbolya, cczhou, fyxiao, yongjaelee}@ucdavis.edu

【ICCV 2019】

摘要：

在 MS COCO 数据集上做出了第一个实时的实例分割模型，在 ones stage 目标检测的基础上增加了一个 mask 预测分支，生成若干个 prototype mask，然后加权组合再裁剪得到实例的 mask。还提出了比 NMS 算法更快的 Fast NMS。

方法：

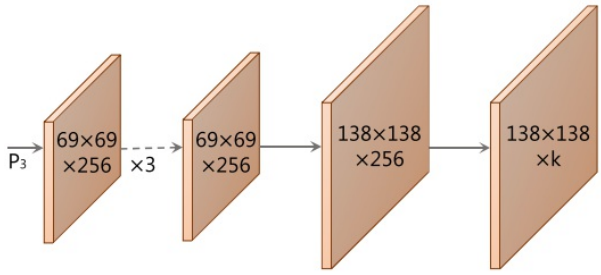


网络的主干使用 resnet101，用 FPN 来多尺度预测，以及生成更加精细的 mask。

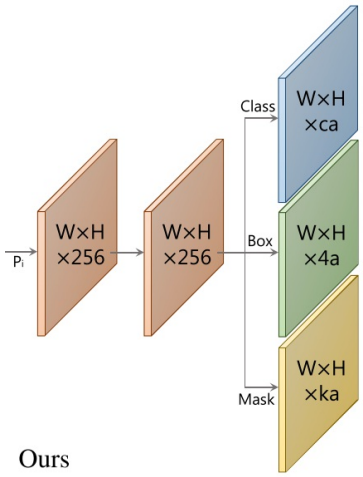
接下来是两个并行的分支，分别生成 k 个 prototype mask，以及目标检测任务，

但是多了 K 个 mask 系数的回归，根据系数将原型 mask 组合，并用目标框裁剪再二值化得到每个实例的 mask。

Protonet 输出 k 个 138*138 的 mask：



Prediction Head 的输入为 P3~P7，输出 c 个类别预测，4 个 bbox 预测值，以及 k 个 mask 组合系数。其中第一层卷积每个分支都共享，对于每个尺度，每个像素点生成 3 个 anchor，比例是 1:1、1:2 和 2:1，五个特征图的 anchor 基本边长分别是 24、48、96、192 和 384。



fast nms：

首先取出每一类的所以 n 个 box，根据目标得分降序排序，计算互相之间的 IoU 得到一个 n*n 的矩阵；

取这个矩阵的上三角矩阵，然后再取这个上三角矩阵的每一列的最大值，一个 n 个值；再对这 n 个值做阈值，大于阈值的 box 将删除。

这样做的原因是，由于每一个元素都是行号小于列号，而序号又是按照置信度从高到低降序排列的，因此任一元素大于阈值，代表着这一列对应的 RoI 与一个比它置信度高的 RoI 过于重叠了，需要将它舍去。

损失函数：

类别置信度的 loss，使用 smooth L1；

位置偏移的 loss，使用 smooth L1；

mask loss，使用的是 pixel-wise 的二分类交叉熵。

总结：

存在问题：定位不准，比 mask rcnn 差的最重要原因。边缘泄露：如果两个目标离得比较远，网络的原型 mask 不会倾向于将他们分开，因为还有裁剪步骤；但是这都依靠与定位准确性。